HA31
T67

QA385
T67

Cornell University Library
HA 31.T67
Steps in the solution of the correlation
3 1924 013 704 071
mann

STEPS IN THE SOLUTION OF THE CORRELATION COEFFICIENT USING RECTANGULAR CORRELATION CHART REVISION, NUMBER 6.

Determine which variable shall be the independent (X-variable) and which the dependent (Y-variable) and write their names into the appropriate boxes at the top and left-hand sides of the page respectively. The independent variable should be that variable which bears a more or less causal relationship to the other variable which is said to be dependent upon it. Thus, if we are correlating intelligence and scholarship, intelligence (the presumable causal variable) should preferably be the X-variable and scholarship (the resultant variable) should be the Y-variable.

Determine the class interval, I_X, and the classes of the X-variable as follows:-
Determine the highest score (H) and the lowest (L) of the X-variable. Subtract the lowest score from the highest score and add 1 to the remainder; this quantity is the Inclusive Range and indicates the possible number of different gross scores to be taken into account in our classes. Divide the inclusive range by 20 and raise the quotient to the next higher integer. This quotient is the class interval, I_X, which should be entered in the I_X-box at the top of the page. Now, in that compartment of the X-classification which is labeled "Lowest X-class" enter the beginning or zero-class of the X-distribution. Begin all the classes with a multiple of the class interval, I_X. Write in all the other classes in ascending order toward the right, Then repeat the X-classes in the horizontal row through the middle of the chart which is labeled "Repeat X-classes in this row."
Caution: Do not omit writing in any classes just because you happen to know that you have no frequencies in a given class, for instance. This is a common error of beginners.
Illustration; the lowest score (L) in a given intelligence test is 8. The highest score (H) is 96. The inclusive range is therefore 96 - 8 + 1 = 89. Now, $\frac{89}{20} = 5$, when we raise the quotient to the next higher whole number. Therefore, the proper class interval is 5. Step-0 of the X-class should now begin with a multiple of the class interval, that is, with a multiple of 5. Since the lowest score is 8, we must therefore begin the 0-step with 5, rather than with 10 or 15, in order that the score of 9 may be included in it. Our steps therefore fun as follows:

Step-0	5-9
Step-1	10-14
Step-2	15-19, etc.

Cornell University Library

The original of this book is in
the Cornell University Library.

There are no known copyright restrictions in
the United States on the use of the text.

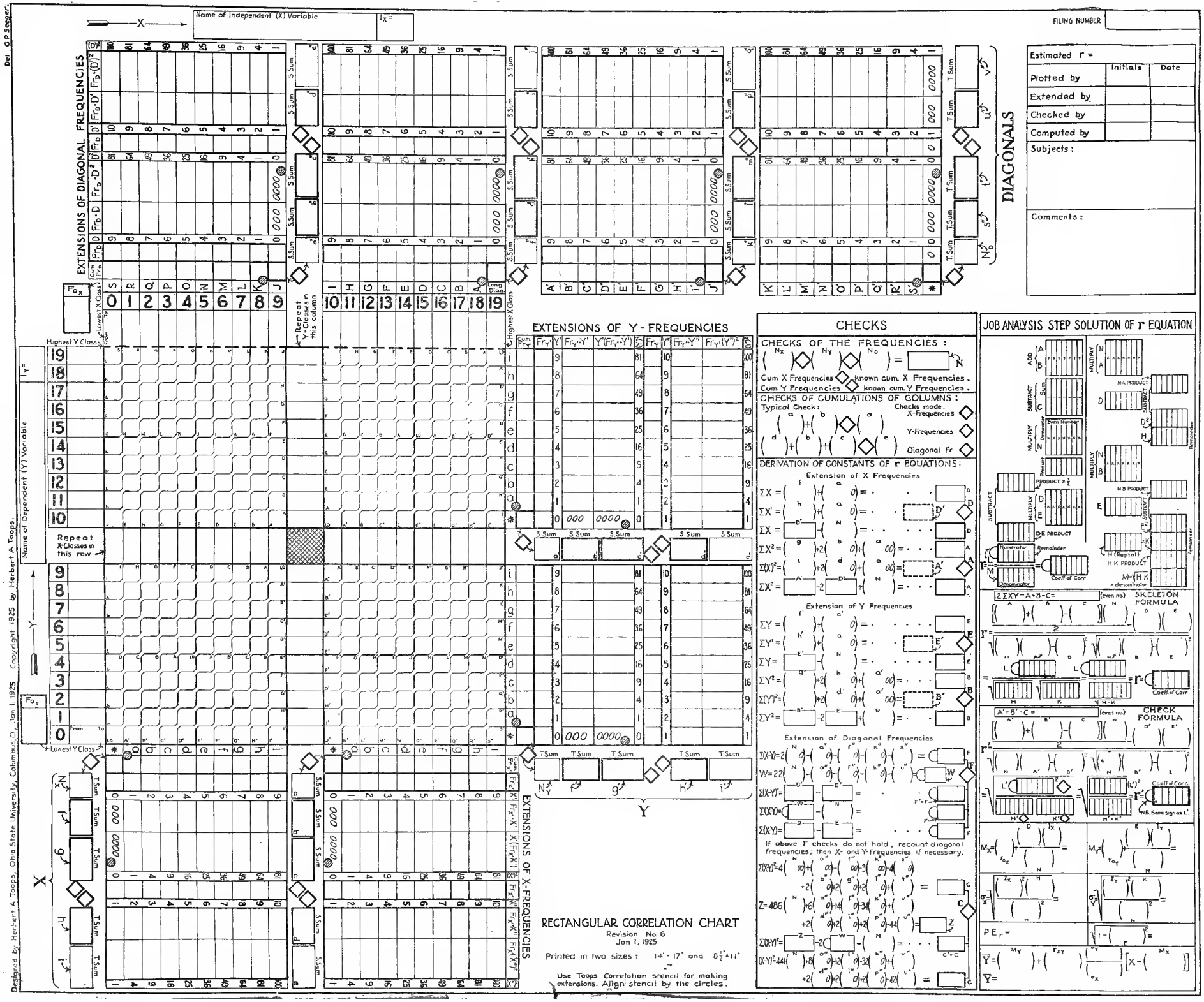
http://www.archive.org/details/cu31924013704071

3. Determine the face value F_{o_x} of the 0-step of the X-variable. This is readily done as follows: Decide what is the lowest possible value, carried out to 4 decimal places, which can be plotted in the 0-step; then decide the very highest value, carried out to 4-decimal places, which can be plotted in the 0-step. Add these two values and divide by 2. The result is the face value of the 0-step.
Illustration: Thus if the 0-step is, 5-9, as above, the conventional usage of statisticians generally is to consider that this includes all scores between 4.5001 and 9.4999 inclusive. Adding these two values gives 14.0000; dividing 14.0000 by 2 gives 7.0000, which is the face value of the X-variable, and 7.0000 should now be entered into the F_{o_x} -box. This technique of determining the face value of a step should be carefully memorized as its use will prevent the very great confusion that is common in interpolating the true mean from a correlation plot.
Caution: Sometimes the tabulated score, 5, does not mean all those values from 4.5001 up to and inclusive of 5.4999. If, for instance, we had taken the ages of children only as on their last birthday then the tabulated age, 5, would mean the age of children who are at least past 5 and not yet 6; consequently, in this case, the lowest value which can be plotted in the class 5 is 5.0001 and the highest value is 5.9999 and F_{o_x} for age 5 in this case is $\frac{5.0001+5.9999}{2} = 5.5000$. By using this value for Fo_x, when using this method, one will obtain the true mean for the children just as accurately as if he had taken the ages of the children to their nearest birthdays. We could of course have taken the ages of the children according to the age which they will be at their next birthday.
In this case Fo_x for the class 5 = $\frac{4.0001+4.9999}{2}$ = 4.5.*
In all work with tests it is desirable to assume that a test score, 5, extends from 4.5001 up to 5.4999 inclusive, in order that we shall conform with conventional statistical usage generally, and in order that any average obtained from grouping the data will check with the average found by the customary method of: (a) adding up all the gross scores and (b) dividing the aggregate thus obtained by the number of cases.

4. Repeat procedures 2 and 3 for the dependent, or Y-variable, repeating the Y-classes in the column headed "Repeat Y-classes in this column."

* This last usage is seldom employed.

Missing Page



Name of Independent (X) Variable
FILING NUMBER
Estimated r =
Initials
Date
Plotted by
Extended by
Checked by
Computed by
Subjects :
Comments :
EXTENSIONS OF DIAGONAL FREQUENCIES
DIAGONALS
EXTENSIONS OF Y-FREQUENCIES
EXTENSIONS OF X-FREQUENCIES
Highest Y Class
Lowest Y Class
Name of Dependent (Y) Variable
Repeat Y-Classes in this column
Repeat X-Classes in this row
CHECKS
CHECKS OF THE FREQUENCIES :
Cum X Frequencies known cum. X Frequencies.
Cum. Y Frequencies known cum. Y Frequencies.
CHECKS OF CUMULATIONS OF COLUMNS :
Typical Check :
Checks made.
X-Frequencies
Y-Frequencies
Diagonal Fr
DERIVATION OF CONSTANTS OF r EQUATIONS :
Extension of X Frequencies
Extension of Y Frequencies
Extension of Diagonal Frequencies
If above F checks do not hold, recount diagonal frequencies; then X- and Y-frequencies if necessary.
JOB ANALYSIS STEP SOLUTION OF r EQUATION
SKELETON FORMULA
CHECK FORMULA
Coeff. of Corr.
N.B. Same sign as L'.
RECTANGULAR CORRELATION CHART
Revision No. 6
Jan 1, 1925
Printed in two sizes : 14" × 17" and 8½" × 11"
Use Toops Correlation stencil for making extensions. Align stencil by the circles.
Designed by Herbert A Toops, Ohio State University, Columbus, O., Jan 1, 1925 Copyright 1925 by Herbert A Toops.
Del G P Seeger

5. Plot the paired measurements as check marks in the compartments. When plotted it is desirable to enter the compartment frequencies as a small figure in the lower corner of each compartment. In using the size 8½" by 11" chart, where the compartments are small, it may be desirable to enter these compartment frequencies only, and after they have been obtained by some other method of tabulation as by use of a Hollerith Sorting Machine or by tabulation on large sized cross-section paper.

6. Add the vertical frequencies by columns, and enter the sums at the foot of the proper columns in the column headed, Fr_X'.
Cautions: Do not enter the frequencies in the column headed, Cum.Fr_X'. When adding the frequencies, fill up any empty boxes in the Fr_X, Fr_Y' and the Fr_D columns respectively with dashes, to indicate that you have considered the frequencies in the particular row and have decided that the row has 0-frequency. This simple caution alone will prevent many errors in solving correlations. Enter the numbers with their vertical axes in a N.E.- S.W. direction; i.e., in such a position that they may be readily read when looking at the N_X, f, g, h, i, windows.

7. Add the horizontal frequencies by rows, entering the sums at the right-hand ends of the proper rows in the column headed Fr_Y'.

8. Now take the frequencies in the Fr_X column and add them downward from the top of the column toward the bottom of the page; when reaching the a-box take off a sub-sum and enter in the a-box; then continue adding on to that sum the frequencies of those steps below the a-box until the N-box is reached when we enter the total number of cases in the N_X-box. If N_X does not check with the known number of cases one must discover the error at once and correct it; the error will probably be found to consist in a wrong addition, although it is barely possible that a frequency may not have been tabulated or one may have been tabulated twice. In order to determine whether one frequency has not been tabulated, add up the Fr_Y column in exactly similar fashion, writing in the sub-sum at a' and the total sum at N_Y. If N_X and N_Y do not check with each other or with the known number of cases, the error is usually readily found and corrected.

9. Cumulate the X-frequencies from the top of the column toward the foot, that is, toward the N_X-box. Write the first cumulative frequency directly opposite that frequency which is nearest the top, entering it in the column headed "Cum. Fr_X';" then add on each successive frequency toward the bottom of the chart until all the cumulative frequencies are obtained. When computing intercorrelations one may check the cumulative frequency column with the standard or "correct" cumulative frequency column and note any errors. If, upon applying this check, there are any errors which are not due to addition in the cumulative column, it is necessary to correct the plotting and the sums of rows and columns, and then to correct any entries previously made in the a and N_X windows, and in the cumulative windows. Compare the cumulative frequency in the two bold-faced cumulative frequency boxes with the numbers previously recorded in boxes a and N_X. If these check place a check mark in the two little check diamonds, with the double arrows, provided for that purpose.

Caution: In case any of the X-steps have 0-frequency, the cumulative frequency of that step is the same as the preceding cumulative frequency; that is, 0 added on to a previous cumulative frequency does not change the cumulative frequency. Again, if there are no frequencies in the 0-step of the variable, it should be noted that the cumulative frequency for this box is, of course, N.

10. Aided by the little lanes, and using a paper or celluloid stencil or serrated edge, or preferably by drawing lightly pencilled lines through the diagonal lanes, add up the diagonal frequencies and enter the sums in the row at the top of the page in the column headed Fr_D. When adding up the long diagonal (LD) enter the sum of the long diagonal frequencies in that compartment of the Fr_D row which is immediately above the label "Long Diag." when adding up the frequencies in the diagonal row marked "A" enter the sum in the Fr_D row just above the label "A" at the top of the page; similarly when adding up the A'-diagonal row write the sum above A' at the top of the page; etc. Notice that the S-diagonal row has only one compartment in it; the R-diagonal row has 2 compartments in it; the Q-diagonal row has 3 compartments, etc.
Caution: Note that the A-diagonal row consists of 3 different sections in as many different quarters of the table; and so similarly for other diagonal rows from A-I inclusive except the LD-diagonal row, which has only 2 sections. Diagonal rows, J to S, have sections in one quarter only of the table.
Do not overlook any of those diagonal frequencies: the safest way to insure no frequencies being overlooked is to draw a lightly pencilled line with a pencil of medium hardness through each diagonal row as you count its frequencies. In counting the diagonals, if it should be found that any diagonal row has a 0-frequency we should not fail to place a dash in the appropriate entry box in the Fr_D-row in order to prevent the possibility of later on shifting of the succeeding entries one unit too far to the left or to the right as the case may be. We should remember that the check of the frequencies adding up to N is a check only to the fact that we have the right total number of frequencies in the column and not on the fact that these frequencies are in the correct boxes; that the checks on the extensions are checks only on whether the extensions have been correctly done for those frequencies placed in the boxes in which they have been placed and that these two checks alone will hold if the arithmetical work has been done correctly irrespective of whether the diagonal frequencies are entered in the correct compartments of the Fr_D- row or not. This pencilled line should not be heavy enough to in any way interfere with reading the frequencies of the compartments which the line crosses. Take the rows in order, first the A-row, then the B-row, the C-row, and so on.

11. Add the Fr_D-frequencies from the top of the page toward the bottom, that is, toward N_D. Enter dashes in all compartments of the Fr_D-row which do not have any frequencies in them. Take off sub-total sums at the a", f", and k"-windows, and finally enter the total sum in the N_D window.
Caution: If there are no frequencies in that section of the table which is above a", then the a"-entry is 0. If there should be no entries in that section of the table which is below the k"-window then the entry for the k"-window is the same as N_D, the total number of cases.

Check N_D with N_X and N_Y. If this check holds we may now enter check marks in the first four check diamonds of the "CHECKS" section of the sheet and enter the value of N in the rectangular box marked "N" in this section.

12. Cumulate the frequencies from the top of the page toward the bottom, that is, toward N_D, taking account of the cautions given under step-9 above. If the cumulative frequencies in the bold-faced boxes of the Cum. Fr_D -row check with the frequencies in the a", f", k", and N_D-windows respectively we may now enter check marks in the four check diamonds, with the double arrows, provided for the purpose.
Note that there is no standard cumulative frequency with which we can compare this column such as we have for the cumulations of the X and Y variables when computing inter-correlations when we have available standard or "correct" cumulative frequencies for each variable.

13. Now recopy the frequencies of the Fr_X, Fr_Y, and Fr_D, columns into the columns of the same headings respectively, to the right in each of the three sections of the extensions of the table. Take care to make your figures legible.

14. Align the window of the correlation stencil by placing the little circular hole so as to coincide with the little cross-sectioned circle provided on the chart for that purpose, thus making the *window of the chart coincide with the *window of the stencil; the a-row of the chart align with the a-row of the stencil; etc.
<u>Caution</u>: Do not confuse the legends *, a,b,c,d,e,f,g,h,i,j,etc. here used for aligning the stencil with the a,b,c,etc.,a',b',c', etc. windows in the body and at the foot of the respective columns of the <u>extensions</u>. The stencil can be used only on the 14" by 17" chart. In the case of the $8\frac{1}{2}$" by 11" chart the extensions can usually be done very readily by mental multiplication since the multipliers never are larger than 10. On the stencil, in that stencil column which corresponds to the frequency of the a-frequency-compartment of the chart, find in the a-row of the stencil the $Fr_{X'}$. X' and X' ($Fr_{X'}$.X') entries; the former in light-faced type and the latter in bold-faced type, and enter these on the chart in the proper columns. Continue thus for all the frequencies of the X-variable. Do both the top and bottom halves or sections of the table.

15. Move the stencil to the right and align the stencil again by means of the little cross-sectioned circle. Enter the extensions $Fr_X'.X''$ and $Fr_X'(X'')^2$ in the appropriate places. Do both sections of the table. On the $8\frac{1}{2}$" by 11" chart these extensions cannot be done by stencil but must be done mentally using as multipliers the X", $(X'')^2$, multipliers given on the chart

16. Carry out procedures 14 and 15 for the Y-variable and also for the diagonal frequencies. All of these can thus best be done at one time before the additions of the extensions are begun.

17. Beginning at the top of the columns and adding toward the bottoms, that is toward N_X, add all the columns of the extensions of the X-frequencies. Using an adding machine we can readily use the machine to add two columns at once by entering the $Fr_X'.X'$ extensions in the left-hand half of the given board and the $X'.(Fr_X'.X')$ extensions in the right-hand half of the key board. When we get to the b and c windows in these two columns respectively we must note the sub-sums and record them accordingly in the b and c windows; then continue on to the foot of the page and we obtain the total sum of the two columns respectively which are entered in the f and g windows. In like manner also we obtain all other sub-sums and total sums in the extensions of the four columns of the X-variable, the Y-variable and the diagonal frequencies.
<u>Caution:</u> If in the sections of the extensions which are above the b, b' and b" windows there are no entries except 0-then the sub-sums at these points will of course be 0. This applies to all such similar situations. At this stage the person who made the extensions of the frequencies by use of the correlation stencil, signs his initials in the upper right hand corner in the row "Extended by"

18. The checker of the extensions now begins his work using the typical checks provided in the "CHECKS" section of the table. The sum of the a+b windows in the X-variable must check with the entry in the d window; if so a check mark may be placed in the first check diamond found between windows c and d. Then the sum of windows d+b+c should check with the entry in window e. If this check holds a check mark may be placed in the second check diamond found between windows c and d. If these checks do not hold the error <u>in making the extensions</u> must be found and corrected. If the checks do hold all the extensions in the upper half of the extensions of the X variable are correct. Now apply the same checks to the N_X, f, g, h and i windows. Since in the place of the a-boxes we have here the N_X-box, and in place of the b-box we have the f-box etc., it is necessary to make these appropriate substitutions. It is simpler to think of the checks in this section as;-

The first two boxes N+f must check with the third box to the right, namely, with h: then as the second check, the three middle or bold-faced boxes (f+g+h) must check with the fifth box, or i. If these checks do not hold and the previous checks have been made to hold by correcting any errors that were found, then the errors will be found in the bottom half of the extensions and can readily be located by noting which of the two checks fails to hold. If the second check does not hold then the error will be found in the extensions for the g-or the c-windows. If both of these checks hold, then all of the work, on the X-frequencies, both in making the extensions and in adding them, is correct.
Caution: Do not forget that the g-window is the sum of the c-window plus all the extensions intervening between the c- and g-windows. This error, if made, would have occurred in the previous step.

19. In similar manner check all the extensions of the Y-frequencies and all the diagonal frequencies. These all having been checked as correct, the checker signs his initials and date at the upper right hand corner after the words "Checked by."
Caution: If an error in the extensions has occurred in the top section of the extension column, say in b, this will of necessity introduce an error into the total sum of the column, in f. In order to save time the total sum of the column in the f-window should be corrected accordingly before applying the check to the bottom half of the column. In the diagonals an error in the b" sum will be reflected also as an error in the g", i", and s"- windows all of which will need to be corrected.
Note: Some people prefer to make the check at the very bottom of each column. If this check holds, then the checks above that point do not need to be made. If this bottom check does not hold then one can perform the previous checks and thereby very quickly locate that section of the table which contains the errors.

20. At this point the computor enters into the windows of that section of the table which is headed "Derivation of Constants of r Equation" the values therein required. Note that in that part of this section which is headed "Extension of X-Frequencies", the quantities to be entered will come from the general portion of the extensions of the table which is marked off by the long brace labeled "X". Similarly for the extensions of the Y-frequencies, and the extensions of the diagonal frequencies. However, the entries D', E', D and E cannot be entered until computed in the two previous sections. Having entered all the possible values in the respective parentheses of this "CHECKS"- section, the computor proceeds to blindly follow out the computations, [illegible]

Note that the formula for [illegible] in reality calls for f + 10a, and that in order to simplify the problem, 0 has been placed in the (a) in order to indicate the multiplier of 10; thus, if f = 86 and a = 2 we read the a as 20 in making computations so that D is accordingly 86+20 = 106
[illegible]

Caution: In case any of the values called for are 0's, dashes should be entered in those boxes respectively rather than 0's, in order to obviate the possible error of calling a true zero in the b window in solving for the ΣX^2, 200, say, instead of its true value. In case b has the value 0, it is even better to draw the dash through the 2 and including both parentheses. In the entry $\Sigma X^2 = (514) + 2(20) + (200)$, we read the sum as follows: $514 + 2(20) + 200$ which is 754. That is, the original formula, for which this stands, is $\Sigma X^2 = g + 20b + 100a$. But, we must not fail to note that, in order to save time and reduce the complexity of the multiplication, the 0 of the 20 and the two 0's of the 100 are placed within the b and the a-parentheses respectively. As soon as D has been derived twice independently in the two proper boxes the two values should check. If they do, we may place a check mark in the D check diamond; similarly having obtained the second value for A independently, if it checks with the first value we may place a check mark in the A check diamond, likewise for E and B respectively.

The quantity F is derived four times independently in the section of the table headed "Extension of Diagonal Frequencies." In case of certain errors it may happen that two of these F-values will check but that these will fail to check with the remaining values. If all four values of F, derived independently, check, then we are certain that the counting of the diagonals has been done correctly, and that they have been entered in the proper windows in the Fr_D-row. If the four F-checks do not all check out to the same value, then one should first reassure himself that the arithmetic in this section of the checks has been done correctly; then after no error is detected in the arithmetic, the diagonal frequencies must be recounted in order to note particularly whether some diagonal frequency has not been placed one box too far to the right or to the left. Such an error seldom occurs in case we have taken the precautions under step-10 above, of always entering a dash in the Fr_D-row whenever a diagonal row has 0-frequency. Once in a great while, when solving isolated correlation coefficients where there are no standard cumulating frequencies for X and Y as in intercorrelations we may have to recount the X-and Y-frequencies. Such an error will be very much more infrequent than the errors in counting the diagonal frequencies.

The work being correct to this point, one proceeds to determine the value of C by three independent methods. If the first two values of C check the third value will always check unless an arithmetical error has been made in determining the third value for C. Consequently the third independent derivation of C may or may not be made, depending upon one's time.

In all the above a calculating machine will be found to be by far the most efficient method of solving the "CHECK"-formulae.

21. The various checked values above found are now entered in the "Skeleton Formula" and the "Check Formula" and the two equations are solved for r. Note that C and N in both formulae are the same value but that A' is always larger than A; B' is always larger than B, etc. A calculating machine is the most ready method for simplifying the two equations, resulting in a reduction down to three integral numbers. At this point, L should check with L'. If so, a check mark may be placed in the check diamond at the right of L'. Similarly H' must check with H, and K' with K whereupon check marks may be placed in the check diamonds provided. If these checks do not hold, note whether the substitutions of numbers for letters has been correctly made in the two formulae; then recalculate that one or ones of the three quantities L, H, and K, which do not check.

In case one wishes to do so, and does not have a calculating machine available, he may make use of the "Job Analysis Step Solution of r Equation" given at the top of the column.* The reduction is performed merely by following out in systematic order from the top downward the directions, "Add, Subtract," etc. The little +, −, and × signs also assist in recalling the operation which is to be done at any particular point.

22. The final solution for r, after finding the reductions L, H, K, may be found by slide rule, which is the quickest method, or by performing numerically the operations indicated on the sheet, and in the order indicated.

(a) Solution of r by slide rule.

Set the hair-line over the numerical value of L on the D-scale of the slide rule; bring the value of H of the B-scale under the hair-line; move the hair-line to the index of the C-scale; bring the numerical value of K on the B-scale under the hair-line; opposite the index of the C-scale, find on the D-scale the answer sought. This may be checked by the following procedure: Bring the hair-line opposite L on the D-scale; now bring K on the B-scale under the hair-line; slide the hair-line to the index of the C-scale; bring H on scale B under the hair-line; read r opposite index of scale C on scale D.

(b) Solution of r by logarithms.

Or it may be solved by logarithms as follows: Find the logarithm of H and also of K; add the two and take half of the sum; subtract this half-sum from the logarithm of L; look up the anti-logarithm of the remainder. This anti-logarithm is r.

(c) Solution of r on the calculating machine.

On the calculating machine solve r in two different ways. (A) in the skeleton formula, multiply H by K and extract the square root. Divide this square root into L which gives r as the quotient. (B) Square L'; multiply H' by K'; Divide $(L')^2$ by H' and K'; and extract the square root. The result is r. However the sign of r must be given the same sign as L'.

Caution: Do not forget to give the negative sign to r in case L' has a negative sign.

Note: In all the above reductions, as well as in the derivation of the constants of the r equations any quantity which may be negative has a semi-circle attached on its left; such quantities do not come out negative in all problems but occasionally may do so.

* The vertical lines are intended to separate the successive digits of each number and so accurately to align them in the column.

23. <u>The solution of the averages</u>. Substitute the numerical values in the formula for M_X and M_Y AND solve. The result gives the means directly without any interpolation. We have previously shown in step 3 above how to find Fo_X and Fo_Y. As a check on the solutions it is desirable to solve them again by the following interpolation formula: $M'_X = \frac{D}{N}$; $M'_Y = \frac{E}{N}$. If the result of this operation is, say $M_X = 8.2$, this is to be interpreted as the face value of the 8th step plus two-tenths of a class interval of the X-variable. The face value of the eighth step of the X-variable can be readily found; to this we add I_X (.2). The result must check readily with M_X solved by the skeleton formula of the chart.

24. The solution of σ_X and σ_Y. In these solutions the quantities H and K have been found above and have been checked in the solution of the skeleton formula and the check formula for r. Note that the standard deviation is always positive. As a check on the solution of σ_X one may solve the formula in the following manner:
(a) extract the square root of H
(b) divide this square root by N
(c) multiply the quotient by I_X.

25. The $P.E._r$ and the coefficient of alienation, K, may be copied into the chart from prepared tables or may be solved by the usual formulae:

$$P.E._r = \frac{1 - r^2}{\sqrt{N}} \qquad K = \sqrt{1 - r^2}$$

It will be noted that the numerator of the $P.E._r$ upon extraction of the square root becomes K.

26. <u>The solution of the regression equation for predicting</u> Y, (the dependent variable). Substitute the values of M_Y, r_{XY}, σ_Y, σ_X, M_X in the formula of the chart and solve. The final equation is always of the form $Y = a + b \cdot X$. For a check on the equation, solve the following equations:

$$b = \frac{L \cdot I_Y}{H \cdot I_X} \qquad a = Fo_Y + I_Y \left(\frac{H.E - D.L}{NH}\right) - \frac{L.I_Y.Fo_X}{H.\ I_X}$$

Since this formula involves all of the quantities used in finding the correlation coefficient, standard deviations, and the means, if this formula is carried out to three decimal places and checks with the previous formula the result is a check on the correctness of the solution of the five quantities involved, as well as upon the correctness of the regression equation. If the check does not hold, and if the arithmetic of the above two formulae has been done correctly, then it is necessary to examine the standard deviations for the errors involved in them.

Copyright 1922, by Herbert A. Toops

Filing Number

X →

I_X =

I_Y =

Y ↓

Y-Classes

X-Classes

Fr (X-Y)²	Fr (X-Y)²																														
																			0000												
(X-Y)²	289	256	225	196	169	144	121	100	81	64	49	36	25	16	9	4	1	0		1	4	9	16	25	36	49	64	81	100	121	144	169	196	225	256	289
	Q	P	O	N	M	L	K	J	I	H	G	F	E	D	C	B	A	Longest Diagonal		A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q

Q	Q	P	O	N	M	L	K	J	I	H	G	F	E	D	C	B	A	C	q		17			289
P	P																	A	p		16			256
O	O																	B	o		15			225
N	N																	C	n		14			196
M	M																	D	m		13			169
L	L																	E	l		12			144
K	K																	F	k		11			121
J	J																	G	j		10			100
I	I																	H	i		9			81
H	H																	I	h		8			64
G	G																	J	g		7			49
F	F																	K	f		6			36
E	E																	L	e		5			25
D	D																	M	d		4			16
C	C																	N	c		3			9
B	B																	O	b		2			4
A	A																	P	a		1			1
From	To																	Q	*		0	000	0000	0
		A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q						

X-Classes	From	To	*	a	b	c	d	e	f	g	h	i	j	k	l	m	n	o	p	q
				0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
				000																	
				0000																	
				0	1	4	9	16	25	36	49	64	81	100	121	144	169	196	225	256	289

Fr_Y | Y | $Fr_Y \cdot Y$ | $Y(Fr_Y Y)$ | Y^2

Sum N | Sum E | Sum B

Sum C | Sum N

Sum N | Sum D | Sum A

Job Analysis Step Solution of r Equation

Add A, B

Subtract Sum C

Multiply Remainder N

Product

Product × ½

Multiply D, E

Subtract

D·E Product

Multiply N, A

N·A Product

Subtract D

D^2

Remainder H

Multiply N, B

N·B Product

Subtract E

E^2

Remainder K

H (repeated)

H·K

$r = \frac{L}{M} =$ Coeff. of Corr.

L = Remainder − Remainder

M = Denominator

$M = \sqrt{H \cdot K}$

Skeleton Formula –

$$r = \frac{[(A) + (B) - (C)](N) - (D)(E)}{2\sqrt{(N \cdot A) - (D)^2}\sqrt{(N \cdot B) - (E)^2}} = \frac{L}{\sqrt{H}\sqrt{K}} = r = \text{Coeff. of Corr.}$$

$M_X = \left(\frac{D}{N}\right) =$ Steps — Interpolation M_X =

$M_Y = \left(\frac{E}{N}\right) =$ Steps — Interpolation M_Y =

$\sigma_X = (I_X) \cdot \sqrt{\left(\frac{H}{N}\right)^2} =$

$\sigma_Y = (I_Y) \cdot \sqrt{\left(\frac{K}{N}\right)^2} =$

RECTANGULAR CORRELATION SHEET

Designed by Herbert A. Toops, Institute of Educational Research, Teachers College.

© Herbert A. Toops

11. If a copy of the stencil, "Correlation Extension Tables", is not available, the small numbers placed near each row or column for multiplication by the frequencies, and labeled X, X^2, Y, Y^2, $(X-Y)^2$, may be used as multipliers of the frequencies instead of using the stencil products. The stencil greatly hastens this operation and tends to eliminate the chances for error. These small numbers take the place of the usual deviation steps from an assumed mean, and are always positive, thus no negative products are produced.

12. Using only the bold-faced products of the stencil above, find the $Fr_D \cdot (X-Y)^2$ products corresponding to the respective diagonal frequencies, Q, P, O, etc. Check by multiplying Fr_D by $(X-Y)^2$ multipliers of chart.

13. Add all the quantities in the five rows and columns from which are obtained the sums which are entered respectively in the D, A, E, B, and C windows.

14. Copy the window sums into the proper lettered places in the skeleton formula and solve for r. A calculating machine is very efficient in reducing the formula to the final form, involving the reductions L, H, and K, which may be solved efficiently by slide rule.

15. If desired, operation 14 may be replaced by copying instead the window entries into the lettered windows of the "Job Analysis Step Solution of Equation for r". The reduction is performed merely by following out in systematic order from the top downward the directions, "Add, Subtract", etc., as found. The final solution of r is by slide rule as before.

16. To find the X-average, or M_X, substitute the proper window quantities in the skeleton formula and solve. The answer is in terms of table steps; thus if, M_X equals 4.6, this must be interpreted to be the face value of the d-column, (class No. 4 in the X-variable) plus six-tenths of the class interval, I_X. The average of Y is similarly found. As a check on the averages solve the following formulae:

$$M_X = \frac{D}{N} \cdot I_X + Fo_X \cdot M_Y = \frac{E}{N} \cdot I_Y + Fo_Y,$$

where Fo_X and Fo_Y are the face values of the zero-step of the X-variable and the Y-variable respectively.

17. To find the standard deviation of X, or σ_X, substitute the proper window quantities in the skeleton formula and solve. Note that H is found in the reduction of the denominator of r. The standard deviation of Y is found in similar fashion.

18. The window sums, in order are:

N = N, number of cases	$A = \Sigma X^2$	$B = \Sigma Y^2$
$D = \Sigma X$	$E = \Sigma Y$	$C = \Sigma (X-Y)^2$

Reference: Toops, H. A., - Eliminating the Pitfalls in Solving Correlation: A Printed Correlation Form. Journal of Exp. Psych. Vol. 4, No. 6, Dec. 1921, pp. 434-446.

19. CHECKS. - It is desirable to have a check on the accuracy of the entering of the diagonal frequencies. There are three checks:

A. The diagonal frequencies must add up to N.

B. A certain quantity F must check with D-E. Find the quantity, F, by three steps as follows:

(1) Multiply all diagonal frequencies to the left of the "Longest Diagonal" by the square-root of $(X-Y)^2$. In the reference cited above, page 435, this quantity is 11(1)+18(2)+21(3)+35(4)+14(5)+15(6)+6(7)+0(8)+1(9)= 461. Since these Fr_D (X-Y) products are negative, give this sum a negative sign, or -461.

(2) Multiply all diagonal frequencies to the right of the "Longest Diagonal" by the square-root of $(X-Y)^2$. This sum, in the reference cited, page 435, is 7(1)+1(2)= 9. This sum is positive, since it comes from Fr_D (X-Y) products which are all positive.

(3) To obtain F, add algebraically the two sums just found. Thus in the above example, F= -461+9= -452. F must now check with D-E. In the above example D=607; E=1059; F=D - E, or -452 = 607 - 1059 = -452. If this check does not hold, re-check entries of Fr_D frequencies, and then of Fr_X . X and Fr_Y . Y extensions until the error is found.

C. Finally in the solution of the formula for r, A+B - C is always an even number (one divisible exactly by 2) if the extensions have been correctly done. All of the above three checks must hold on any sheet correctly computed.

PRICES: Correlation sheets, each 4¢; stencils, each 10¢; all prepaid. Directions free. Address, Herbert A. Toops, Ohio State University, Columbus, Ohio.